Simultaneous position, scale, and rotation invariant pattern classification using third-order neural networks

Max B. Reid, Lilly Spirkovska, and Ellen Ochoa

Intelligent Systems Technology Branch, NASA Ames Research Center, Moffett Field, CA 94035, USA

Abstract: *We demonstrate a third-order neural network that distinguishes between classes of patterns regardless of their translational position, scale, and angular orientation. A significant feature of this network is that it is trained on only one view of each pattern, using a simple single-layer perceptron learning rule. In approximately one minute of run time on a Sun 3 computer, the network learns to distinguish between the letters T and C at any position, scale, or rotation in a 9 x 9 image field, with 100% accuracy in a noise-free background. Examples of both second-order and third-order networks illustrate that geometric invariances can be built into the network architecture using information about the relationships expected between input pixels. The invariances achieved require no learning to produce and apply to any input pattern learned by the network. Higher-order neural networks are therefore capable of efficiently performing both types of mapping required by pattern recognition problems, namely feature extraction and object classification.*

1. Introduction

Pattern recognition may be viewed as a two part process of feature extraction followed by object classification [1–4]. First, a preliminary mapping from an image to a representation space is made, generally resulting in a significant degree of data reduction. A second mapping then operates on this reduced data to produce a classification or estimation in an interpretation space. Historically, these steps have required either mathematical mappings operating directly on a detected image [1,2] or initial feature extraction performed through optical processing followed by some form of analytical discrimination.[3]

Both mappings may also be performed using neural network models [4]. In this paper we discuss neural networks both as classifiers in hybrid systems and as implementations of the complete pattern recognition operation. Emphasis is given to recognition invariant to distortions in scale, translational position and angular orientation. The relatively poor results with neural models performing the complete mapping from image to interpretation is attributable to the unsuitability of the models used for distortion invariant feature extraction. In contrast, higher-order neural networks can be designed to implement the extraction of simple but effective features suitable for in-plane distortion invariance. Simulation results of higher-order neural networks demonstrating simultaneous invariance to scale, translation and rotation will be presented.

2. Neural networks for pattern recognition

Pattern recognition requires the nonlinear separation of pattern space into subsets representing the objects to be identified. Early research into neural networks concentrated on defining their potential for nonlinear discrimination [5,6]. It was found that a single layer, first-order neural network can only perform linear discrimination. However, either multilayer, first-order networks or single layer networks of higher order can provide the desired nonlinear separation [6].

The capability of neural networks to perform nonlinear separation can be applied both to extract image features and to interpret images based on a feature set. Practical applications in distortion invariant pattern recognition have been found for hybrid systems utilizing neural networks for classification. Troxel *et al* [7] successfully applied a multi-layer perceptron neural network trained with a backward error propagation (back-propagation) learning algorithm [8,9] to classify laser radar images of targets, invariant to position, rotation and scale. The data was first mapped into the magnitude of the Fourier transform with log radial and angle axis, $|F(\ln r, \theta)|$, feature space. Glover [10] describes a practical product-inspection system based on the optical Fourier transform and neural classification. Rotation and scale invariance has also been described in a system using complex-log conformal mapping combined with a distributed neural associative memory [11]. In all of these approaches utilizing neural classification, distortion invariance is achieved through non-neural feature extraction techniques.

It has been argued that nonlinear neural computing is theoretically superior to methods such as matched filters or linear correlation for the complete pattern recognition operation, including feature extraction [12]. However, the performance of neural networks to date fails to fulfill this promise. For instance, several types of neural associative memories have been shown to be computationally more expensive than matched filters in a study involving the recognition of line segments [13]. Multi-layer networks trained by back-propa-

gation have also been applied to recognition tasks, examples being sonar signal classification [14] and distortion invariant character recognition [15,16]. In these cases, the networks achieved ≈80–90% recognition accurancy only after being shown a training set of images several hundred [14] or thousand [15,16] times. Learning by back-propagation to distinguish a 'T' from a 'C', invariant to translation and rotation, required over 5000 presentations of an exhaustive training set [15]. Learning to distinguish 36 patterns in a 5 x 5 pixel array invariant to translation required over a 1000 training set presentations to a network composed of two-layers, each with 25 Adelines arranged in slabs [16].

The relatively poor performance of neural networks in the preceding examples, most particularly the failure to produce efficient distortion invariant recognition, is due to the fact that first-order networks are poorly suited for extracting distortion invariant features. One layer of a typical first-order network is shown in Figure 1.

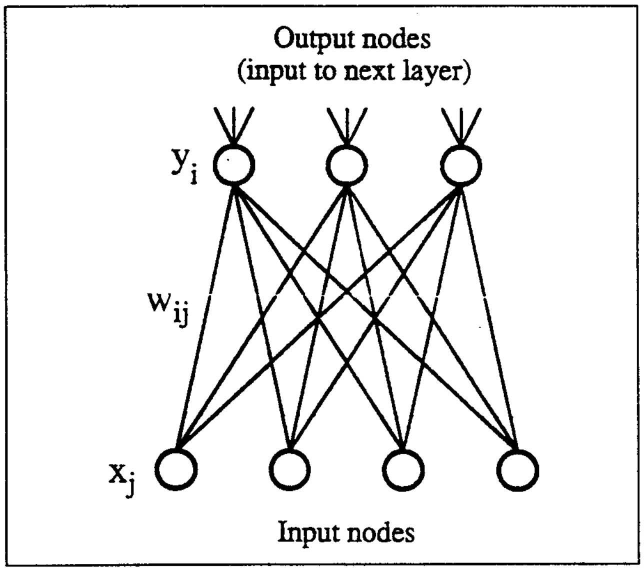

Figure 1: One layer of a first-order neural network.

The activation level of an output node in a first-order neural network is determined by an equation of the form:

$$y_i = \Theta(\Sigma_j w_{ij} x_j) \tag{1}$$

where Θ is a nonlinear threshold function, the x_j are the excitation values of the input nodes, and the interconnection matrix elements, w_{ij}, determine the weight that each input is given in the summation.

Achieving translation, scale and rotation invariance requires a neural network to learn relationships between the input pixels, x_j . Note that the summation within the parenthesis in Eq. (1) is a function of individual x_j's. No advantage is taken of any known relationships between the x_j's. Multilayer, first-order networks can learn invariances, but require a great deal of training, and produce solutions that are specific to particular training sets.

A further disadvantage is that the mappings learned are opaque: it is not readily evident what features are being extracted or how classification is being performed. It is generally assumed that the output of intermediate-layer hidden nodes in the network correspond to specific features, and in some applications it is possible to discern what these features are [14]. In distortion invariant recognition application, however, it is not apparent that first order networks' hidden nodes come to represent efficient feature sets or even feature sets sufficient to allow classification by succeeding layers.

3. Higher-order neural networks

The output of nodes in a general higher order network is given by:

$$y_i = \Theta(\Sigma_j w_{ij} x_j + \Sigma_j \Sigma_k w_{ijk} x_j x_k + \Sigma_j \Sigma_k \Sigma_l w_{ijkl} x_j x_k x_l + ...) \tag{2}$$

A diagram of a neural network utilizing only second-order terms is shown in Figure 2. Higher-order neural networks (HONNs) were evaluated in the 1960s for performing nonlinear discrimination but were rejected as impractical due to the combinatoric explosion of higher-order terms [6].

Recent research [17–19] has shown that the problem of combinatoric explosion can be overcome by building invariances into the network architecture using information about the relationships expected between the input x_j's. HONNs are thus well suited for invariant pattern recognition because feature extraction is functionally built into the architecture. The invariances achieved require no learning to produce and apply to any input pattern learned by the network. Further, a HONN can perform nonlinear discrimination using only a single layer so that a simple perceptron learning rule can be used, leading to rapid convergence [4].

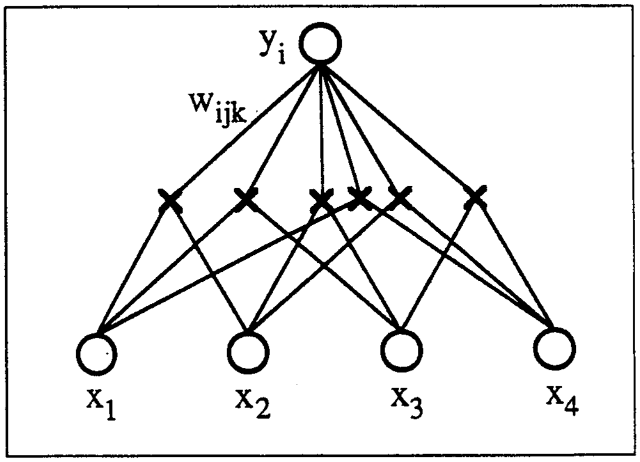

Figure 2: A second-order neural network with 4 inputs and 1 output.

As an example, translation invariance can be built into the second-order neural network with 4 input nodes and 1 output node shown in Figure 2. Assume that the input patterns (1 0 1 0) and (0 1 0 1) are to be identified as the same object. If $w_{i13} = w_{i24}$ then y_i is the same for both inputs. In general, translation invariance requires that:

$$w_{ijk} = w_{i(j-k)} \tag{3}$$

i.e., the connections for equally spaced input pairs are all set equal.

Combinations of invariances can similarly be achieved. A second-order neural network will be simultaneously invariant to scale and translation if the weights are set according to the function [18]

$$w(i,j,k) = w(i,(y_k - y_j)/(x_k - x_j)) \tag{4}$$

Equation (4) implies that w_{ijk} is set equal to $w_{ij'k'}$ if the slope of a line drawn between nodes j and k equals that formed between j' and k', as shown in Figure 3.

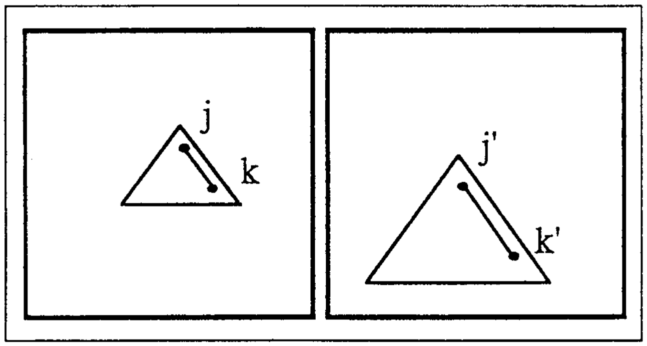

Figure 3: Translation and scale invariance achieved by setting $w_{ijk} = w_{ij'k'}$ if the slope of the line formed by nodes j and k equals that formed by nodes j' and k'.

Any object drawn in a 2–D plane can have lines of various slopes drawn within it. An object's relative content of lines of different slopes does not change when it is translated in position or scaled in size, as long as it is not rotated.

Rotational invariance can be included by using a third-order neural network, where the output is given by the function

$$y_i = \Theta(\Sigma_j \Sigma_k \Sigma_l w_{ijkl} x_j x_k x_l) \tag{5}$$

As shown in Figure 4, any three points within an object define a triangle with included angles (α ,β ,γ). When the object is translated, scaled and rotated, the three points in the same relative positions on the object still form the included angles (α ,β ,γ). Therefore, invariances to all three distortions can be achieved with a third-order network having an interconnection function of the form:

$$w_{ijkl} = w_{i\,\alpha\beta\gamma} = w_{i\,\gamma\alpha\beta} = w_{i\beta\gamma\alpha} \tag{6}$$

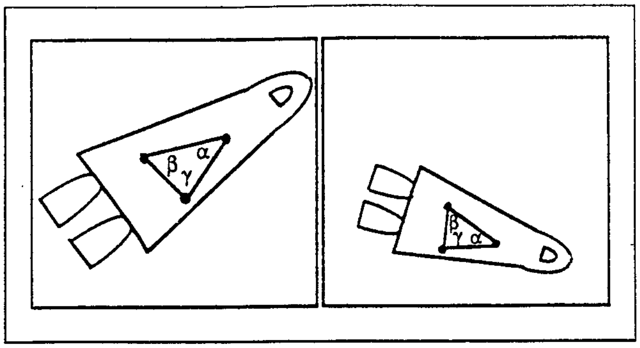

Figure 4: Translation, scale and rotation invariance is achieved by setting all third order weights equal for sets of inputs j, k, and l which form similar triangles.

Note that the order of angles matters, but not which angle is measured first.

4. Simulation results

We have simulated both second- and third-order neural networks to achieve simultaneous invariance to (1) translation and scale with a second-order network, and (2) translation, scale and in-plane rotation using a third-order network. The single layer, second-order network is simulated using a 16 x 16, or 256 node, input field fully interconnected to a single output note which is thresholded with a fixed-threshold hard limiter:

$$\Theta(\Sigma) = 1, \text{ if } \Sigma > 0,$$
$$\Theta(\Sigma) = 0, \text{ if } \Sigma \leq 0 \tag{7}$$

There are 256-choose-2 or 32,640 input pairs and therefore interconnections. The interconnection weights are constrained to follow Eq. (4) in order to achieve invariance to scale and translation. The weights are initially set to zero and a perceptron learning rule is used:

$$\Delta w_{ijk} = (t_i - y_i)\, x_j x_k \tag{8}$$

where the expected training output, t, actual output, y, and inputs x, are all binary. The network is trained on just two distinct patterns — only one size and one location for each pattern. It learns to distinguish between the patterns after approximately ten passes of the training set, requiring less than one minute of run time on a Sun 3 workstation. After training, it successfully distinguishes between all translated and scaled versions of the two objects with 100% accuracy. No further training is required to achieve this invariance, as it is built into the architecture. The system can learn to distinguish between any two distinct patterns, and has been tested on a variety of problems, including the T-C problem [5]. Scale invariance of a factor of 5 has been demonstrated for this problem, with 100% recognition accuracy.

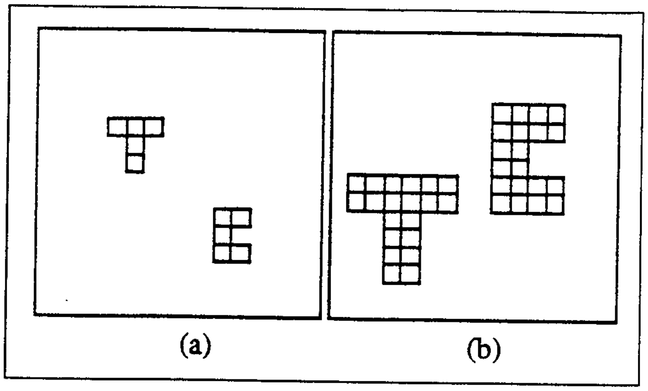

Figure 5: Two different scales of T and C drawn in a 16 x 16 pixel window.

Due to the limited resolution of the finite 16 x 16 input window, residual scale variance can occur. (T,C) pairs are distinguished by their relative content of horizontal and vertical information. For the smallest (T,C) pair, shown in Figure 5a, the T has three input pair combinations arranged horizontally and three vertically, while the C has two arranged horizontally and four vertically. In the next larger scale of (T,C), shown in Figure 5b, the ratio of horizontal to vertical pixel pairs is 34:34 for the T and 26:42 for the C. It is therefore easier to distinguish between the smalled (T,C) pair based on their relative horizontal/vertical content. If the system is trained on the smaller set of letters, learning is not pushed to the point where larger versions can be recognized. In contrast, if large patterns are used for training, all smaller versions are subsequently recognized.

Residual scale variance can be eliminated by using bipolar training values and a modified threshold function such as,

$$\begin{aligned} &\Theta(\Sigma) = 1, \text{if } \Sigma > K, \\ &\Theta(\Sigma) = -1, \text{if } \Sigma < -K, \\ &\Theta(\Sigma) = 0, \text{otherwise}, \end{aligned} \qquad (9)$$

where K is some positive constant. Learning with a sufficiently large value for K forces the network to make a greater distinction between the initial patterns, allowing easier discrimination between test patterns which are subsequently evaluated with a hard limiter. Training the network on the smallest (T,C) pair using a value of K = 1000 allows correct identification of all larger test versions, without greatly increasing the training time.

For the third-order network simulation an input window of 9 x 9 pixels, or 81 input nodes, is used. The 81-choose-3, or 85,320, weights are constrained to follow Eq. (6) in order to achieve invariance to scale, translation, and in-plane rotation. The weights are initially set to zero and a learning rule is used of the form:

$$\Delta w_{ijkl} = (t_i - y_i)\, x_j\, x_k\, x_l \qquad (10)$$

The training set consists of two images, one for each object to be learned. After approximately 20 passes through the training set, representing ≈1 minute of run time on the Sun 3, the network learns to distinguish between distortions of the two objects with 100% accuracy. The T-C problem can be learned, as shown in Figure 6, with full invariance to translation within the input field, to scale over a factor of three, and to 90° rotations. In principle, recognition is invariant for any rotation angle, given sufficient resolution to draw the image accurately at arbitrary angles.

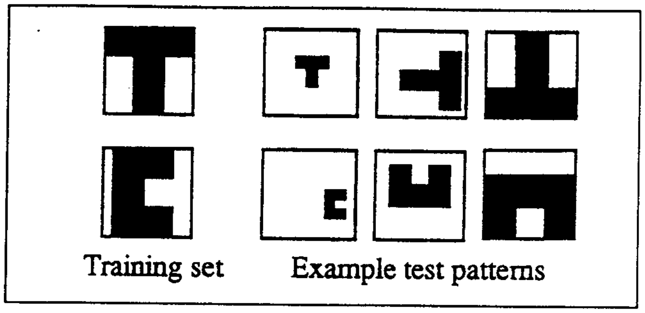

Figure 6: Training set and sample test patterns for distinguishing a 'T' and a 'C', invariant to translation, scale, and rotation.

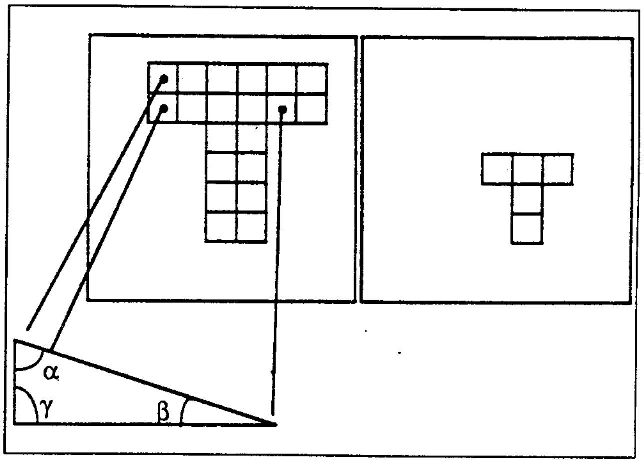

Figure 7: A triangle with included angles, α, β, and γ which may be drawn between the pixels of a 6 x 6 pixel 'T', but not between those of a 3 x 3 pixel 'T'.

As in the case of the second-order network, the small window size leads to some residual scale variance. The triangles which can be formed between the pixels of the smallest T or C vary considerably from those which may be formed with larger versions of the letters. Figure 7 shows an example of a triangle which included angles α, β, and γ formed by three pixels of a 6 x 6 pixel T. These angles are not enclosed by any triangle which can be drawn on the smallest, 3 x 3 pixel, T. In this case, residual scale variance is eliminated by decreasing the resolution to which the angles α, β, and γ in Eq. (6) and Figures 4 and 7 are calculated. With larger window sizes, both the image resolution and the resolution to which α, β, and γ are calculated can be increased.

5. Conclusion

Our simulations have demonstrated that a second-order neural network can be rapidly trained to distinguish between two patterns regardless of their size and translational position. 100% recognition accuracy is achieved for several different training pattern pairs using a 16 x 16 input field size. Additional invariance to in-plane rotation has been achieved using a 9 x 9 input field. In both cases, training requires only 10–20 presentations of just one example of each object to be learned. Comparing these results in terms of recognition accuracy and learning speed show HONNs to be vastly superior to multilayer first-order networks trained by back-propagation for this application.

This superiority results from the HONN architecture's ability to perform simple, transparent feature extraction. These simple features, slopes between input pixel pairs in the case of the second-order network, and included angles between input pixel triplets for the third-order network, are sufficient to allow the network to rapidly learn to classify patterns. The provision of a transparent feature extraction mechanism allows a HONN to efficiently perform the complete mapping from image to intermediate feature space to interpretation space required for distortion invariant pattern recognition.

References

[1] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

[2] C.H. Chen, *Statistical Pattern Recognition*, Hayden, 1973.

[3] Q. Tian, Y. Fainman, Z.H. Gu and S.H. Lee, 'Comparison of Pattern Recognition Algorithms for Hybrid Processing', *J. Optical Soc. of America A.*,**5**, 1988, pp. 1655–1669.

[4] Y.H. Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, 1989.

[5] F. Rosenblatt, *Principles of Neurodynamics*, Spartan, 1962.

[6] M.L. Minsky and S. Papert, *Perceptrons*, MIT Press, 1969.

[7] S.E. Troxel, S.K. Rogers, and M. Kabrisky, 'The Use of Neural Networks in PSRI Target Recognition', *Proc. IEEE Int. Conf. on Neural Networks*, San Diego, California, July 24–27 1988, Vol. 1, pp. 593–600.

[8] P. Werbos, 'Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences', Ph.D. thesis, Harvard University, 1974 (unpublished).

[9] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, 'Learning Internal Representations by Error Propagation', *Parallel Distributed Processing*, Vol. 1, Ch. 8, MIT Press, 1986.

[10] D.E. Glover, 'An Optical Fourier/Electronic Neurocomputer Automated Inspection System', *Proc, IEEE Int. Conf. on Neural Networks*, San Diego, California, July 24–27 1988, Vol. 1, pp. 569–576.

[11] H. Wechsler and G.L. Zimmerman, 'Invariant Object Recognition Using a Distributed Associative Memory', *Neural Information Processing Systems*, American Institute of Physics Conference Proceedings, 1988, pp. 830–839.

[12] H. Szu, 'Three Layers of Vector Outer Product Neural Networks for Optical Pattern Recognition', *Optical and Hybrid Computing*, SPIE Vol. 634, 1986, pp. 312–330.

[13] P.M. Grant and J.P. Sage, 'A Comparison of Neural Network and Matched Filter Processing for Detecting Lines in Images', *Neural Networks for Computing*, American Institute of Physics Conference Proceedings, 1986, pp. 194–199.

[14] R.P. Gorman and T.J. Sejnowski, 'Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets', *Neural Networks*, **l**, 1, 1988, pp. 75–89.

[15] D.E. Rumelhart, *op. cit.*, pp. 348–352.

[16] B. Widrow and R. Winter, 'Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition', *IEEE Computer Magazine*, **21**, March 1988, pp. 25–39.

[17] G.L. Giles and T. Maxwell, 'Learning, invariance, and generalization in high-order neural networks', *Applied Optics*, **26**, 1987, pp. 2972–4978.

[18] G.L. Giles, R.D. Griffin, and T. Maxwell, 'Encoding geometric invariances in higher-order neural networks', *Neural Information Processing Systems*. American Institute of Physics Conference Proceedings, 1988, pp. 301–309.

[19] M.B. Reid, L. Spirkovska, and E. Ochoa, 'Rapid Training of Higher-Order Neural Networks for Invariant Pattern Recognition', *Proc. Joint Int. Conf. on Neural Networks*, Washington, D.C., June 18–22, 1989.

The authors

Max Reid

Max Reid received his BS in Electrical Engineering and Computer Science from the University of Colorado, Boulder in 1983, and his MS and PhD in Electrical Engineering from Stanford University in 1986 and 1988 respectively. Dr Reid is currently the Photonics Group Leader in the Intelligent Systems Technology Branch at NASA Ames Research Center. His research includes optical and neural pattern recognition techniques and optical implementations of neural networks. He is a member of the IEEE, the Optical Society of America, the American Institute of Physics, and the International Neural Network Society. He has published 15 journal and conference articles in the areas of inter-actions of high-energy electron beams with electromagnetic radiation, neural networks and optical processors.

Lilly Spirkowska

Lilly Spirkowska received her BS in Mathamatics and Computer Science from the University of Denver in 1984, and her MS in Computer Science from the University of California at Berkeley in 1986. Ms Spirkowska's research in the Intelligent Systems Technology Branch at NASA Ames Research Center emphasises neural pattern recognition techniques. She is also interested in machine learning and intelligent user interfaces. She is a member of the Association for Computer Machinery.

Ellen Ochoa

Ellen Ochoa is Chief of the Intelligent Systems Technology Branch at NASA Ames Research Center, functioning as the technical and administrative head of a staff of 40 engaged in research and advanced development of space-based intelligent computational systems. Before coming to Ames, she spent three years as a staff researcher in the Imaging Technology Branch at Sandia National Laboratories in Livermore, working on distortion- invariant pattern recognition and optical morphological transforms. She received her BS in Physics from San Diego State University in 1980, and her MSEE and PhD from Stanford in 1982 and 1985 respectively. Dr Ochoa has published 15 journal and conference articles in the areas of photorefractive crystals, nonlinear optical processing and optical pattern recognition. She is a member of the Optical Society of America, SPIE — the International Society for Optical Engineering, and the International Neural Network Society.